This Page Is Inserted by IFW Operations and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):



BLACK BORDERS

- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.



11) Publication number:

0 079 739

A2

12

EUROPEAN PATENT APPLICATION

(21) Application number: 82305926.6

(22) Date of filing: 08.11.82

(51) Int. Cl.³: C 12 N 15/00

C 12 N 1/00, C 12 P 21/02 C 07 H 21/04, C 07 C 103/52 //C12R1/19, C12R1/865

(30) Priority: 12.11.81 US 320632

Date of publication of application: 25.05.83 Bulletin 83/21

Designated Contracting States:
BE CH DE FR GB IT LI NL SE

(1) Applicant: THE UPJOHN COMPANY 301 Henrietta Street Kalamazoo, Michigan 49001(US)

(72) Inventor: Dugaiczyk, Achilles c/o The Upjohn Company 301 Henrietta Street Kalamazoo Michigan 49001(US)

(2) Representative: Perry, Robert Edward et al, GILL JENNINGS & EVERY 53-64 Chancery Lane London WC2A 1HN(GB)

(S) Albumin-based nucleotides, their replication and use, and plasmids for use therein.

(5) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

EP 0 079 739 A

UPJOHN

٠ :

GJE 70/2056/02

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in

5 development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly α-fetoprotein, but the synthesis decreases drastically after birth. Recently,

10 Law et al determined the complete sequence of mouse α-fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been 15 reached from studies on the α-fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al,

J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum 20 mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA comprises 2,078 nucleotides, template. The sequence starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-serleu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino 35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the tripledomain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION-

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleo-25 tides, of which 38 represent the 5'-untranslated region, 54 identify a prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table 1) are complementary to a 3'-terminal region of eukaryotic 185 RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T
$$T^{C}T$$
 C T T C T G T......albumin mRNA (3')...G A G G A A G G C G U C C $m_{2}^{6}A$ $m_{2}^{6}A$185 RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

4083

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since prepeptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a prepeptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the propeptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the concensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

TABLE 1

	(30)	(170)	(190)	(350)	(440)	(330)	(029)	(710)	(300)
5	AGC AGC	20 1ys AAA	S		110 Pro CCA	140 177	170 gln ÇAA	200 cys TCT	230 atu GAA
	phe TTT	phe TTC	phe TTT	80 thr leu ACT CTT	asn AAC	1eu	169 cys TG:	lys AAG	ata GCA
	Jes CTC	AAT	a) CAA	879 GCA	aso GAC	tyr TAC	168 cys TCT	leu CTC	phe TTT
	phe leu TTT CTC	£ 8	thr	val	asp	1 × 8	5 ¥	ar AGA	alu GAG
		ala GA	val GTA	AC th	lys AAA	1ys AA		aln CAG	ala CCT
10	-10 leu leu phe leu CTY CTY TTY CTC	9 2	glu val GAA GTA	75 lys leu cys thr AAA TTA TGC ACA	. 1 2	phe leu lys lys tyr TTT TTG AAA AAA TAC	ala phe thr GCI TTI ACA	1ys AA	phe pro lys ala glu phe TTT CCC AAA GCT GAG TTT
		leu 77G		leu TTA	ain his CAA CAC	phe TTT	ala	3] a	i S
	11e ATT	asp	val GTG	1 ys	leu ain his TTG CAA CAC	th ACA	ala GCT	ser TCT	phe TTT
	phe TTT		leu TTA	asp CAC	Phe	a)u GAG	1ys AAA	ser TCG	arg AGA
	p r o trp val tlu phe lle ser TGG GTA ACC TTT ATT TCC	phe	40 qlu asp hís val lys leu val asn GAA GAT CAT GTA AAA TTA GTG AAT	70 phe qly TTT CGA	101 938 760	a ja GA	tyr TAT	ala GCT	a]n CAG
15	val t	10 87.9 CCG	40 481 GTA		100 alu GAA	130 asn C AAT	160 arg AGG	190 1ys AG	220 Ser AGC
	trp TGG	his CAT	his CAT	thr leu ACC CTT	asn	ase GAC	1ys	a) y 666	leu CTG
	1ys AAG	ala his GCT CAT	asp his GAT CAT	thr	2 2	M is CAT		alu GA	arg 000
•	-18 Met ATG	glu val GAG GTT	phe glu TTT GAA	h is CAT	pro aly ara CCT GGG AGA	ala phe GCT TTT	ohe TTT	arg asp CGG GAT	val ala GTA GCT
20	-18 Met GCTTTTCTCTTCTGTCAACCCCACAGCCCTTTGGCACA ATG	9 fc 6 QC	phe TTT	Lea CTT	P70	ala GCT	phe TTC	8 19 000	220 ala phe lys ala trp ala val ala arg leu ser GCT TTC AAA GCA TGG GCA GTA GCT CGC CTG AGC
20	7507	ser AGT	pro CCA	36T	gin glu CAA GAA	thr	leu CTT	glu leu GAA CTT	trp ala TGG GCA
	icc11	lys ser AAG AGT	34 gin gin cys pro CAG CAG TGT CCA	lys AAA	gla	124 asp val met cy8 thr GAT GTG ATG TGC ACT	leu CTC		tr 166
	וכאפנ	ala his GCA CAC	ule CAG	62 cys asp TGT GAC	ala lys GCA AAA	met ATG	₽6. 85	asp GAT	818 GCA
	200		gla CAG	62 cys TGT	ala GCA	val GTG	pro CCG	leu CTC	7. AA
25	CAAC	1 asp GAT		60 glu asn GAA AAT	91 cy3 TGT		ala ೧೧೧	lys AG	phe
	יכדפו	pre 53	30 TAT		90 cys TGC	120 glu val GAG GTT	150 phe tyr TTT TAT	180 pro lys leu asp CCA AAG CTC GAT	210 arg ala AGA GCT
	יוכדו	arg CGT	913	ala GCT	asp GAC	glu GAG	phe TTT	Jec 77G	gly glu arg GGA GAA AGA
	1110	p r o val phe GTG TTT	ala cct	3er TCA	ala GCT	£ 55	tyr TAC	leu CTG	gly glu GGA GAA
	8		phe TTT	g)u GAG	met. ATG	arg AGA	pro CCT	973 160	91y CGA
30		-6 arg oly AGG GGT	နှင့် တွင	asp GAT	glu GAA	val GTG	413 CA T	ီ နှင့် ၁၁၁	afg TT
		-6 arg AGG		ala GCT	aly GGT	1eu	5 S	ala GCT	17 A
		← ₽ S	16 11 G	val CTT	tyr TAT	879 CGA	8 79 AGA	1ys	ŧ ₹
		tyr IAT	21 ala leu val leu lle GCC TTG GTG TTG ATT	55 57 107 107 107 107 107 107 107 107 107 10			\$ 1 \$	asp	ser leu gin lys AGT CTC CAA AAA
25		ala CCT	leu TTG	A t	glu thr GAA ACC	leu pro CTC CCC	11e ATT	ala GCT	ser AGT
35		ser ala tyr a TCG GCT TAT T	21 ala GCC	51 lys thr AAA ACA		asn AAC	161 glu 11e GAA ATT	171 ala CCT	201 ala GCC

	(890)	(980)	(1070)	(1140)	(1250)	(1340)	(1630)	(1520)	(1610)	(1700)
r	260 leu CTT	290 11e ATT	320 ala GrT	350 ala GCC	380 Jeu CCT	410 ara CGT	840 h1s CAT	&70 ser AGT	Sno 1ys AAA	530 val GTT
5	ase GAC	289 cys TGC	asn tyr AAC TAT	le CTT	pro CCT	val GTT	lys AAA	val CTA	ورو	leu CTT
	هاه 900	h is CAC		8 A	lys AAA	ala leu leu val CCG CTG TTA GTT	438 cys lys TGT AAA	460 461 leu cys val leu his giu lys thr pro val TTA TGT GTG TTG CAT GAG AAA ACG CCA GTA	val GTT	lys lys ain thr ais leu AAG AAA CAA ACT GCA CTT
	arg	ser	316 cys lys tgr AAA	leu CTG	phe TTT	leu CTG	437 cys TGT	thr	tyr val	thr ala ACT GCA
	asp GAC	1ys	316 cys TGC	leu CTG	glu	ala ACG	Tys AA	17,8	ACA th	e cy
10	35D GAT	glu	asp val CAT GTT	leu CTG	phe asp TTC GAT	asn	aly ser GGC AGC	ale GAG	₽. ₹	lys lys AAG AAA
	ala Cr	leu TTG		val	phe TTC			5 5	asp GAT	
	253 cys TGT	280 glu lys pro leu leu glu lys ser gaa aaa cct ctg ttg gaa aaa tcc	glu ser lys GAA AGT AAG	vál GTC	lys val AAA GTG	lys phe AAA TTC	430 arg asn leu qly lys val AGA AAC CTA GGA AAA GTG	leu 11G	glu val GAA GTC	ain ile CAA ATC
	253 leu glu cys CTT GAA TGT	pro CCT	ser AGT		1ys AA		lys AAA	val GTG	ole CAA	t S
•	250 leu leu CTG CTT	lys AAA		340 asp tyr GAT TAC	370 tyr ala TAT GCC	400 alu tyr GAG TAC	430 leu aly CTA GGA	461 eys TGT	490 ala leu GCT CTG	8 70 A
15	250 leu CTG		310 val	340 88p GAT	370 tyr 1AT		430 1ev CTA	\$60 1eu 17A	490 ala CCT	520 q1u GAG
	asp CAT	279 cys TGT	phe	pro CCT	369 alu cys GAA TGC	leu aly CTT GGA	AAC	o Po CAG	Ser TCA	qlu lys GAG AAG
	his gly	278 279 qlu cys cys GAA TGC TGT	SSP	arg his pro AGG CAT CCT		aln 1eu CAG CTT	arg AGA	AAC	phe TT	ser qlu lys TCT GAG AAG
		5 ₹9	ala GCT	arg AGG	pro his CCT CAT	alu aln GAG CAG	ser TCA	Jeu CTG	cys 160	
20	345 246 cys cys TCC TCC	ser lys leu lys Agt aaa CTG AAG	ala CCT	ala arq GCA AGA		a la GAG	glu val GAG GTC	val val GTG GTC	arg pro cys phe CCA CCA TGC TTT	518 cys thr leu TGC ACA CTT
20	345 cys 100	leu CTG	leu TTA		asp GAT	phe TTT		رهر 100	క్ట్ ప్ర	thr AGA
	245 thr glu cys ACG GAA TGC	ser lys AGT AAA	ser TCA	glu tyr GAA TAT	ala GCA	leu CTT	leu val CTT GTA	3e. 77	asn arg AAC AGG	514 11e cys ATA TGC
			pro CCT		ala GCT	alu GAG		leu CTA		11e
	h is CAC	367 700	leu 11G	tyr TAT	ala occ	392 cys TGT	pro thr CCA ACT	tyr TAT	leu val TTG GTG	asp
25	val GTC	270 ser 11e TCG ATC	asp GAC	1eu 77G	361 cys 1GT	asn	PT0 CCA	asp GAC	leu 11G	510 his ala CAT GCA
	240 173 AA		300 ala GCT	330 phe TTT	360 cys 1CC	390 91n CAA	420 thr ACT	450 glu GAA	\$80 36r TCC	510 h1s CAT
	thr ACC	asp	pro CCT	met ATG	glu lys GAG AAG	IIe lys ATC AAA	ser TCA	448 cys ala TGT GCA	94 6¥3	phe TTC
	le. CTT	gln	met ATG	91y CCC		11e ATC	val		th VO	thr
	asp GAT	glu asn GAA AAT	glu GAG	phe leu TTC TTG	leu CTA	leu TTA	pro gln CCC CAA	met pro ATG CCC	476 477 cys cys TGC TGC	phe TTC
30	val thr GTG ACA		asp GAT		thr	gin asn leu CAG AAT TTA	579 CCC	met ATG	476 cys TGC	thr
	va1 GTG	265 11e cys ATC 1GT	asn	val GTC	thr ACC	ale CAG	val	arg ACA	173	₹ §
	lys leu AAG TTA	11e ATC	val glu GTG GAA	asp GAT	91u G&A	pro CCT	1ys	ala lys GCA AAA	thr ACC	ala GCT
	ser lys leu val thr ICC AAG TTA GTG ACA	261 265 ala lys tyr ile cys GCC AAG TAT ATC TGT	glu val glu asn GAA GTG GAA AAT	321 glu ala lys asp val GAG GCA AAG GAT GTC	351 1ys thr tyr glu thr thr leu AAG ACA TAT GAA ACC ACT CTA	381 val glu glu pro gin asn leu GTG GAA GAG CCT CAG AAT TTA	1ys AAG	441 pro glu ala lys arg met pro CCT GAA GCA AAA AGA ATG CCC	val	501. giu phe asn ala giu thr phe GAG TIT AAT GCT GAA ACA TIC
25	ser TCC	1ys AAG	g lu GAA	ala GCA	thr		th Acc	35 €	arg AGA	phe TTT
35	231 val GTT	261 ala CCC	291 ala GCC	321 glu GAG	351 1ys AAG	381 val GTG	411 tyr TAC	441 pro	471 asp GAC	501 91u GAG

5	558 559 560 1 cys cys 1ys 3 TGC TGF AAG (1790)	ter (CACATTAAAAG (1883)	AAATTETTTAA (2M2)
10	540 pro lys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys cys lys CCC AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC TGF AAG	567 570 cys phe ala glu glu gly lys lys leu val ala ala ser gln ala ala leu gly leu ter ter 70c TTT GCC GAG GAG GGT AAA AAA CTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (1883)	ter ter Catctcagcctaccatgagaataagaaaaatgaagatcaaaagcttattcatgtgttttgttttggtgtaaagccaacagcgtgtgtaaaaagataaattgttaa (2m2)
15	550 (a) met asp asp phe a	580 ila ala ser gin ala a cct GCA AGT CAA GCT G	CTTTTTCGTTGGTGTAAAGC
20	ilu gin leu lys ala v AG CAA CTG AAA GCT (lly lys lys leu val e	AAGCTTATTCATCTGTTTT
25	540 lys ala thr lys g AAG GCA ACA AAA C	570 phe ala glu glu q TTT GCC GAG GAG C	AAGAAATGAAGATCAA
30	531 glu leu val lys his lys pro GAG CTC GTG AAA CAC AAG CCC	567 asp asp lys glu thr cys GAC GAT ANG GAG ACC TGC	ter ter CCTACCATGAGAATAAGAGA
35	531 glu leu GAG CTC	561 ala asp GCT GAC	CATCTCAG

TCATTITGCCTCTTTTCTCTGTGCTTCAATTAAAAAATGGAAAGAATCTAA.... 20AA (2078)

Following are examples which illustrate procedures, including the best mode, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5 Example 1 Isolation of Messenger RNA

٠.

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) <u>Biochemistry</u> 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and Tse [Taylor, J.M. and Tse, T.P.H. (1976) <u>J. Biol. Chem.</u> 251, 7461-7467]. <u>In vitro</u> translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) <u>Nature</u> 227, 680-685.

Example 2 Cloning Procedures

30

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, S., et al., Ibid.]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [\$^{32p}]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

As shown in Example 5, plasmids pHA36 and pHA206 were deposited in <u>E. coli</u> HB101 hosts. The plasmids were obtained from <u>E. coli</u> RR1 hosts, described in this example, and transformed into <u>E. coli</u> HR101 by standard procedures well known to those of ordinary skill in this art. The <u>E. coli</u> RR1 hosts were lysed and then centrifuged to separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris·HCl, pH 8.0, 10 mM CaCl₂, 10 mM MgCl₂). The cells for transformation are

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HB101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl₂. Bacteria are then concentrated to one-tenth of this volume in CaCl₂ and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 μg/ml tetracycline) with 200 μl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Rethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) <u>J. Virol.</u> 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) <u>Biochemistry</u> 11, 1242-1250] gels.

Example 4 DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and γ[^{32p}]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

Example 5 Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable E. coli host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

HB101(pHA206) - NRRL B-12550

5

 $\underline{\text{E. coli}}$ HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public. upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

 $\underline{E.~coli}$ RR1 and $\underline{E.~coli}$ HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEp6 is a well known and widely available yeast episomal plasmid.

20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEp6) - NRRL B-12093.

Example 6 Assembly of the Serum Albumin Gene

Assembling the pieces together is a straighforward task of restriction enzymology. There is only one Mspl site in the overlapping
DNA sequence of the two cDNA clones. Two enzymatic steps of (i) Mspl
digestion of the two DNAs, followed by (ii) the use of ligase, an
enzyme that seals DNA fragments, will give the desired product.
Although two other undesired DNA species will also be obtained in the
course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA
species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

(a) Escherichia coli

- (b) Saccharomyces cerevisiae
- (a) The vector of choice is plasmid pRR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoRl DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one of the yeast plasmid vectors, e.g., YEp6, at the Eco Rl cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the begin-20 ning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been 25 documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466. Example 8 Screening of Clones Producing Albumin

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an <u>in situ</u>

lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

CLAIMS

- 1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.
- Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.
- 3. E. coli HB101 (pHA36) having the deposit accession number 10 NRRL B-12551.
 - 4. $\underline{\text{E. coli}}$ HB101 (pHA206) having the deposit accession number NRRL B-12550.
- 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

	(30)	(170)	(192)	(380)	(660)	(330)	(620)	(710)	(300)
	ACC	23 173 AA	55 818 GCA	8 2 E	110 070 CCA	140 177	170 aln caa (200 cys TGT (230 alu GAA (
5	ohe TTT	ohe TTC	phe TT	thr	AAC	leu 17	169 cys 100	1ys AAG	818 GCA
5	5 CT	AAT	GA A	818 CC	33	tyr TAC	168 cys 1GT	of C	phe TTT
	phe TTT	g Jr GAA	thr Act	נא הדד	asb GAT	1ys AAA	중중	25	alu GAG
	1 E	a ja CA	val GTA	thr	1ys AAA	133 AA	thr ACA	aln ara CAG AGA	s s
	ţ.	93,	£ 8	5 5 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	7.13 CAC	1ec 116	phe TTT	<u>₹</u> ¥	1ys AAA
10		1eu 776	asn		GAA CAA	phe TT	ala GCT	د و در	pro lys CCC AAA
		asp GAT	val GTG	1ys AAA	leu 11G		ala CCT	36 TCT	ohe TTT
		lys AAA	<u>₹</u>	asp GAC	phe TTC	9 9 9	lys AAA	36.	5 A
		phe	1ys AAA	2 5	100 101 alu cys GAA TGC	3 ₹	tyr TAT	190 lys ala AAG GCT	220 leu ser gln CTG AGC CAG
16		10 819 000	40 his val cat gta	70 leu phe CTT TTT	50 ± 50	130 asn AAT	160 arg AGG	190 173 AAG	220 ser gln AGC CAG
15		h1s CAT	his CAT	leu CTT	asn AAT	asp GAC	1 ys	alu aly GAA GGG	arg leu CCC CTG
		ala GCT	asp CAT	th Acc	5 2	his	a a	6 € 6¥	5.50
		val CTT	2₹	MIS CAT	91y 000	phe	ohe TTT	asp GAT	ala GCT
		glu	phe 111	ser leu TCA CTT	pro	ala GCT	phe TTC	919 000	val CTA
20		Ser	pro CCA	ser TCA	g Ju	thr	leu CTT	CTT	e 430
		1ys AAG	34 gin cys CAG TGT	1 ys	e e	124 met cys ATG TGC	leu CTC		ala trp GCA TGG
		h I s	gla	asb GAC	1 ys	met ATG	ale GAA	asp GAT	
		a1a GCA	gln CAG	62 cys TGT	ala CCA	val	970 000	leu CTC	phe 1ys TTC AAA
		asp CAT	Jec CTT	60 glu asn GAA AAT	91 cy3 TGT	asp GAT	a)a ဝင်င	1ys AAG	phe TC
25		- 5 S			80 160	120 val GTT	150 tyr TAT	180 pro	210 818 GCT
		phe arg	gln CAG	ala GCT	SS CAC	g lu GAG	phe TT	leu TTG	5 Y
		r opper	ala CCT	ser TCA	ala CCT	5 Y	tyr TAC	leu CTG	975
		val CTG	ala phe GCC TTT	g Ju GAG	met ATG		5 T	177 cys 100	
30		gly GGT		asp GAT	ag	; arg leu val arg CGA TTG GTG AGA	h is	818 CCC	a pe
30		arg ACG	11e ATT	ala ser	91y CGT	leu 11G	arg AGA	ala GCT	X X
		-1 3er 7CC	1ec 77G	val	tyr TAT	5 43	ala arg CCC AGA	₹ \$	g XX
		-1 ser ala tyr ser rcc GCT TAT TCC	val CTG	53 0ys TGT	th	နို ည	- 8 33	asp GAT	ser leu gln lys phe gly AGT CTC CAA AAA TTT GGA
		ala GCT	lev TTG	51 lys thr AAA ACA	5. €	ie CTC	11e	ala CCT	ser AGT
35	•	36. 700	21 818 GCC	173 AAA	81 819 CCT	111 85n AAC	161 glu 11e GAA ATT	171 ala ala asp GCT GCT GAT	201 ala GCC

	(890)	(980)	(1070)	(1160)	(1250)	(1360)	(1630)	(1520)	(1610)	530 val GTT (1700)
5	260 Jeu CTT	290 11e ATT	320 ala GCT	350 a1a GCC	380 1eu CCT	& 10 8 T G CGT	440 h1s	&70 ser AGT	Sno 1ys AAA (530 val GTT (
	asp GAC	289 cys TGC	tyr TAT	leu CTT	970 CCT	va1 GTT			200	3 5
	260 arg ala asp leu AGG GCG GAC CTT	h I s CAC	asn	ara leu ala AGA CTT GCC	1 X X	14 T	638 cys lys TGT AAA	pro val CCA GTA	val p	2 2
	91.6 ACG	ser TCC		leu CTG	phe lys		637 cys TGT			ر بر د ح
	350 CAC	1 ys	316 ays lys TGC AAA	Jeu CTG	ole:	ala leu occ crc	lys o	lys t	thr tyr ACA TAC	¥ 5 A £
10	asp asp GAT GAC	glu lys GAA AAA		340 asp tyr ser val val leu leu ara leu GAT TAC TCT GTC GTG CTG CTG AGA CTT	SSD	asn i	430 pro thr leu val glu val ser arg asn leu qly lys val qly ser lys CCA ACT CTT GTA GAG GTC TCA AGA AAC CTA GGA AAA GTG GGC AGC AAA	leu his qiu lys thr TTG CAT GAG AAA ACG	alu thr tyr GAA ACA TAC	2.₹
	918 CCT	leu TTG	lys asp val AAG GAT GTT	781	8 5.	gln asn CAG AAT	מן א ממר א	his a	val asp glu GTC GAT GAA	2 S
	253 cys TGT	leu leu CTG TTG	lys AAG	val v	167	phe q	va) q	leu h TG C	81 8 77 G	= 5 ₹.⊀
	253 glu cys ala GAA TGT GCT	pro CCT	ser lys AGT AAG	ser val TCT GTC	73 (AA (2.4	lys v	val 1 GTG T	glu val GAA GTC	- × - ×
	leu CTT	17.8	glu	¥ 5	ala lys val phe asp GCC AAA GTG TTC GAT	tyr lys TAC AAA	7 5 5 7	461 0ys v TCT G	Teu g CTG G	: 5 : 5
15	250 asp leu leu GAT CTG CTT	280 glu lys pro leu leu GAA AAA CCT CTG TTG	310 val CTT	340 pro asp tyr CCT GAT TAC	370 tyr a	400 glu tyr lys GAG TAC AAA	430 leu qly CTA GGA	460 461 ser val val leu asn gln leu oys TCC GTG GTC CTG AAC CAG TTA TGT	490 ala 1 GCT C	520 916 ai GAG A(
	asp	279 cys TCT	310 phe val TTT CTT	. e. c. c.	369 3 cys t TGC 1	AOO aly glu cca cac	asn 1	91n 1 CAG T	ser a	, 50 50 50 50 50 50 50 50 50 50 50 50 50 5
	કે છે		SS CAT	glu tyr ala arg arg his pro GAA TAT GCA AGA AGG CAT CCT	369 ala asp pro his qiu cys GCA GAT CCT CAT GAA TGC	phe qlu qln leu gly TTT GAG CAG CTT GGA	arg a	asu g AAC C	phe s	5 5
	M SAT	278 glu cys CAA TGC	pro ser leu ala ala asp CCT TCA TTA GCT GCT GAT	glu tyr ala arg arg his GAA TAT GCA AGA AGG CAT	h1s c	gln l	ser a	leu a CTG A	cys p	5 G
20	246 233 133		ala CCT	arg AGA	pro P	glu c	al s	val 1 GFC C	pro c	3 5
	240 Jys val his thr glu cys cys AAA GTC CAC ACG GAA TGC TGC	ser lys leu lys AGT AAA CTG AAG	leu ala TTA GCT	818 GCA	asp I	phe c	glu val GAG GTC	ימן י כרכ פ	asn arg arg pi'o cys AAC AGG CGA CCA TGC	ر بر د بر
	5.8		ser TCA	tyr	ala GCA		val 9	ser v TCC	arg arg AGG CGA	514 cys ti TGC A
	th Acc	ser lys AGT AAA	pro	35 8	ala GCT	glu leu GAG CTT	leu v	leu s CTA 7	asn a	7 te 5
	± 50	ser TCC	1ec 11G		\$1\$ 000	392 Cys TGT	thr	tyr 1	val a GTG A	sp 1 AT A
25	240 lys val his AAA GTC CAC		330	Jeu 11G	361 cys TGT	asn c	2 673	asp t	leu val TTG GTG	* &
	240 173 AAA	270 ser TCG	300 ala asp leu GCT GAC TTG	330 phe leu tyr TTT TTG TAT	360 373	390 gln 2	420 thr p	450 glv asp GAA GAC	480 ser 1 TCC T	510 hts a CAT G
	th Acc	gin asp CAA GAT	pro cct		glu lys	¥	ser 1	2 5 5 2 5	glu s GAA T	7
	asp leu thr GAT CTT ACC	gla	met ATG	gly met GGC ATG	ole SAG	. .	al 8	648 oys 8	thr g	ر ج م ۲
20	asp GAT	AAT	glu met GAG ATG	phe leu gly TTC TTG GGC	leu c	gin asn leu il CAG AAT .TTA AT	gin val CAA GTG	5 S		ર ક ક દ
30		glu asn GAA AAT	asp CAT	phe leu TTC TTG	thr 1	15n 1	. 679	met p ATG C	476 477 cys cys TGC TGC	# ₹ 5
	lys leu val thr Ang tta gtg aga	265 cys TCT	88A WT		thr t	gla e	18 P	€ ₹	X	∓ ∓ ≥ ≤
	11A	11e ATC	28	asp v	glu thr GAA ACC	pro g	1ys val AAA GTA	lys arg AAA AGA	thr 1ys ACC AAA	5 G
	lys Ag	tyr lle IAT ATC	781 3TG	lys VG (91c	1ys 1 AAG A	ela 1 CCA A	val ti GTC A	7 B
35	ser lys leu TCC AAG TTA	lys (glu val glu asn asp GAA GTG GAA AAT GAT	ala lys GCA AAG	thr tyr ACA TAT	glv glu pro GAA GAG CCT	thr lys lys val pro gln val ser ACC AAG AAA GTA CCC CAA GTG TCA	648 giu ala lys arg met pro oys ala CAA GCA AAA AGA ATG CCC TGT GCA	arg val AGA GTC	
	231 V81	261 ala lys CCC AAG	291 818 GCC C	321 glu ala lys asp val GAG GCA AAG GAT GTC	351 1ys t AAG A	381 val 9 ctc c	tyr t	441 pro g	471 asp ar	510 glu phe asn ala glu thr phe thr phe his ala asp ile cys thr leu ser qlu lys glu arq qln ile lys lys aln thr ala leu val GAG TET AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TGC ACA CTT TCT GAG AAG GAG AGA CAA ATC AAG AAA CAA ACA GCA CTT GTT GTT

	540 Lys sia thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys cys lys AG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC TGC AAG (1790)	570 phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter TTT GCC GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GQC TTA TAA CATCACATTTAAAAG (1883)	AAGAAATGAAGATCAAAAGCITAITCATCIGITITICTITITCGITGGIGIAAAGCCAACACCCIGICIAAAAAACATAAATITCITITAA (2002)
	60 73 AG () YC (∵ ≨
_	558 559 560 cys cys lys TCC TCC AAG	ter TAAA	III
5	88 88	CATT	ATT
-	1 3 6 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ATCA	ATAA
	7 24	ر اخ	AAAC
	1. Y	eu t	TAAA
10	ž E	28	751
	30 F	3 4) V
	ala a	- 2 22 1	CAAC
	4 E	ala d	MAGG
	550 • 55 CAT	580 gln CA	STGT/
15	\$ 5 T	AGT	GTTC
	met ATG	313 GCA	110
	5.	al &	7.07
	*!s	va1 GTT	1111
20	1 ys	Jeu CTT	TCTG
	leu CTG	1ys AAA	TC
	es CAA	lys AAA	XT7
	9. 86	91y GCT	\ AAG
	₹ ₹	gle Se	¥7.
25	540 thr ACA	570 glu GAG	ž
	₹ 59	عام درر	₩.
	1ys AG	phe TT	Α
	. E 22	567 673 160	20
30	AAG	thr	te.
	± 5	glu	ter ter TGAGAATAA(
	531 glu leu val lys his lys GAG CTC GTG AAA CAC AAG	asp lys glu thr GAT AAG GAG ACC	ter ter CATCTCAGCCTACCATGAGAATAAGAA
	510	asp GAT	3001
	leu CTC	asp GAC	:TCA(
35	syl glu GAG	561 ala CCT	2

TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAAAAATGGAAAGAATCTAA.... 20AA (2078)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

5	(170)	(1560)	(350)	(440)	(330)	(620)	(710)	(300)
	20 1ys AAA	50 818 90	80 CTT	110 070 CCA	140 try TAT	170 gln CAA (200 cys TGT (230 91u GAA (
		phe TTT	thr	AAC	1eu	169 cys 100		sts GCA
	asn phe AAT TTC	5 A	818 GCA	330	tyr TAC	168 cys TCT	leu l CTC /	
10	g Ju GAA	thr		asp GAT	1ys	o le CAA	878 AGA 0	alu phe GAG TTT
10	glu glu GAA GAA	val thr GTA ACT	thr val	1 ys	1 × ×	thr ACA	- u - u - u - u - u - u - u - u - u - u	50 P
	5 5	a A		h Is CAC	leu lys TTG AAA	phe TTT	lys c	lys ala AAA GCT
	asp leu aly glu glu asn phe GAT TTG GGA GAA AAT TTC	asn	75 88p lys leu cys GAC AAA TTA TGC	ain his CAA CAC	130 asp val met cys thr ala phe his asp asn qiu qiu thr phe leu lys lys GAT GTG ATG TGC ACT GTT TTT CAT GAC AAT GAA GAG ACA TTT TTG AAA AAA	ala phe thr GCT TTT ACA	ala iys ain ara leu iys GCC AAA CAG AGA CTC AAG	gin arg phe pro lys ala CAG AGA TIT CCC AAA GCT
		val GTG	1ys AAA	160 11G	A Pr		367 101	phe pro TTT CCC
15	1ys	leu val TTA GTG	asp lys GAC AAA	ohe leu TTC TTG	2 6 6	lys ala AAA GCT	3 e T	ACA 1
	phe TTT	1ys	91y CGA		₹ 8		ala GCT	gln a
	5 5 5 5 5 5	40 val GTA	70 phe TTT	100 101 alu eys GAA TGC	130 asn AAT	160 arg tyr AGG TAT	190 1ys AAG	220 367 ACC (
	#18 CAT	5 Z	leu CTT	AAT	asp GAC	AAA	> C	Jet 0
22	val ala GTT GCT	asp GAT	thr	2 A	hls CAT	ala lys GCT AAA	alu aly GAA GGG	ala arg leu GCT CGC CTG
20	5 TES	₽. Y S	ser leu his TCA CTT CAT	919 GGG	phe	ohe TTT	asp	ala arg GCT CGC
	910	phe TTT	Jeu CTT		thr ala phe ACT GCT TTT	phe TTC	arg 000	val
	ser AGT	Pro	Ser TCA	£ 8	thr	glu leu leu phe GAA CTC CTT TTC	leu	
	1ys AAG	¥ 2°3° 1€1	62 ays asp lys TGT GAC AAA	cAA CAA	124 met cys ATG TGC	leu CTC	glu GA	trp ala TGC GCA
25	his GAC	gla	asp GAC	1ys	met ATG	g fu GAA	asp GAT	a) a
	1 asp ala GAT GCA	gln CAG	62 0ys 1GT	# 55 25	val GTG	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	leu CTC	1 ys
	asp GAT	leu CTT	AAT	91 033 1GT	asp GAT	هاء درد	1ys AAG	210 ala phe GCT TTC
		30 gln tyr CAG TAT	60 91u GAA	98 757 757	120 glu val GAG GTT	150 tyr TAT	180 Pro CCA	210 ala phe GCT TTC
			ala GCT	asp GAC		phe TTT	1eu. 11G	
30		als GCT	ser TCA	ala GCT	5 T T T T T T T T T T T T T T T T T T T	tyr TAC	leu CTG	glu arg GAA AGA
		phe TTT	asp glu GAT GAG	glu met GAA ATG	* 25 \$75	pro		
		313	asp GAT	g g	val arg GTG ACA	his pro CAT CCT	818 000	phe 111
		11e ATT	ala GCT	91y GGT	arg leu val arg CGA TTG GTG AGA	arg AGA	al 8 GCT	1 ys
35		val leu 11e GTG TTG ATT	va! C11	tyr	879 CGA	\$ 5 5	177 Iys ale ale ays AAA GCT GCC TGC	ser leu gin lys phe gly AGT CTC CAA AAA TTT GGA
33		val GTG	53 cys 1GT	thr	leu pro CTC CCC	ele ccc	asp GAT	lei CTC
		leu 11G	tş Ş	35	leu CTC	11e	ala GCT	ser AGT
		21 ala GCC	2 ¥ ¥	B1 arg CCT	111 asn AAC	2 5 8 8	171 a13 CCT	201 ala GCC

5	(890)	(980)	(1070)	(1160)	(1250)	(1380)	(1430)	(0281)	(1610)	(1700)
	260 leu CTT	290 11e ATT	320 ala CCT	350 ala GCC	Jan Teu CCT	ara CCT	880 his cat	470 Ser AGT	500 173 AAA (530 val CTT (
	arg ala asp AGG GCG CAC	289 cys 100	tyr TAT	Jet CTT	Bro CCT	val CTT		val :	bre 1	lea v
	200	MIS CAC	AAC	25	1ys AAA	Jeu TA	438 cys lys TCT AAA	679 (CA (2 P T J .
10		3e. 10	lys AAA	leu leu ara CTG CTG AGA	Phe 1 T T	Jeu CTG		thr a	tyr val	r i
	330 CAC	glu lys GAA AAA	316 cys TGC	leu leu ara leu CTG CTG AGA CTT	glu phe GAA TTT	8 8 335	lys AAA			₹
	253 leu glu cys ala asp asp CTT GAA TGT GCT GAT GAC	pro leu leu glu CCT CTG TTG GAA	val CTT	Jeu CTG	880 GAT	AAT	ser AGC	alu lys GAG AAA	alu thr GAA ACA	8 A 6 G
	ala CCT	Jeu 11G	asb CAT	val GTG	phe TTC	46.0	9. 20.	his c	ass GAT G	ys – AG A
	253 cys TGT	pro leu leu CCT CTG TTG	ser lys AGT AAG	val GTC	val GTG	phe TTC	val GTG	Tro O	val a	TC A
15	28	9.0 CCT	Ser		1ys AAA	lys phe AAA TTC	1ys	val	glu val GAA GTC	1 4 X
	250 leu leu CTG CTT	280 qlu lys GAA AAA	glu	t yr 1 A C	313 SCC		5 2	25.50	leu g CTG C	ខ្
	250 leu CTG		35 5 25 5 17	340 asp cat	370 tyr TAT	400 glu tyr GAG TAC	430 ser arg asm leu qly lys val qly ser lys cys TCA AGA AAC CTA GGA AAA GTG GGC AGC AAA TGT	460 461 leu cys val leu his TTA TGT GTG TTG CAT	ala leu GCT CTG	510 phe his ala asp ile cys thr leu ser qiu iys giu arq qin ile iys iys qin thr ala leu TTC CAT GCA GAT ATA TGC ACA CTT TCT GAG AAG GAG AGA CAA ATC AAG AAA CAA ACT GCA CTT
	asp CAT	278 279 cys cys TGC TGT	phe TTT	870 CC1	369 alu cys GAA TGC		AAC	gla CAG	ser a	7 3 5 7 6 9 5
20	91,	278 glu cys GAA TGC	asp GAT	h is		leu CTT	م م	asn g	phe s	7 7 7 N
	s his	916 CA	ala CCT	glu tyr ala arg arg GAA TAT GCA AGA AGG	4 5 X	ala CAG	ser TCA	leu ser val val leu asn CTA TCC GTG GTC CTG AAC	arg pro cys phe CGA CCA TGC TTT	ici d
	245 246 cys cys TGC TGC	leu lys CTG AAG	ala CCT	arg AGA	pro	alu GAG	val GTC	val GTC	Pro CCA	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
		l leu	3 5	818 GCA	asp GAT	phe .	leu val glu val CTT GTA GAG GTC	va1 GTG	87.8 7.93	514 Lle cys thr leu ATA TGC ACA CTT
	او د ۱۹۶	ser lys AGT AAA	ser TCA	glu tyr GAA TAT	ala GCA	1e CT T	val GTA	ser TCC	5 S	514 cys t TCC /
25	thr Acc		pro CCT	glu GA	ala CCT	58	leu CTT	Jeu CTA	ьво ser leu val asn arg TCC TTG GTG AAC AGG	TA 1
	his CAC	lle ser ATC TCC	140 116	ty.	313 CCC	392 cys TGT	th ACT	7.77	781 s	esp 1
			asp CAC	330 phe leu TTT TTG	361 cys TGT	88n AAT	£ 55	asp GAC	35	P # 70
	240 175	270 3er	300 3.13 CCT		360 cys TCC	390 91n CAA	\$20 thr	450 glu asp GAA GAC	\$6r 700	510 his ala cat gca
20	thr Acc	889 CAT	pro	gly met GGC ATG	glu lys GAG AAG	133 A			9]u	phe trrc c
30	s leu	gla CAA	met ATG		91u 646	11e ATC	val ser GTG TCA	448 oys ale TGT CCA	thr	1. O
	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	265 cys glu asn TGT GAA AAT	glu asn asp glu GAA AAT GAT GAG	leu TTG	leu CTA	Jed TTA			477 273 (glu thr phe thr GAA ACA TTC ACC
	ŞÇ.	26 S	35p	phe TTC	thr	asn AAT		A TG	476 477 cys cys TGC TGC	CA 1
	2 × CT	265 cys TGT	glu asn GAA AAT	val GTC	glu thr thr leu GAA ACC ACT CTA	gln CAG	val GTA	arg met pro AGA ATG CCC	₹ ₹	26.4
35	2 E	11e ATC	g lu GAA	asp CAT	glu GAA		1ys ₩	¥.¥	- L 00	13 CT 0
	. 1 X	ty.	glu val GAA GTG	lys AAG	tyr TAT	glu pro GAG CCT	lys lys	ala lys CCA AAA	al t	asn ala AAT GCT
	36	261 ala lys tyr lle GCC AAG TAT ATC		321 glu ala lys asp val phe leu GAG GCA AAG GAT GTC TTG	351 1ys thr AAG ACA	ole CA	thr.	e nle	476 477 arg val thr lys cys cys AGA GTC ACC AAA TGC TGC	phe a
	231 val GTT	261 ala GCC	291 a1a GCC	321 91u CAG	351 1ys AAG	381 val	tyr thr lys lys val pro gln tac acc aag aad gta ccc caa	441 pro glu ala lys arg met pro CCT GAA GCA AAA AGA ATG CCC	asp a	501 glu phe asn ala glu thr phe thr GAG TTT AAT GCT GAA ACA TTC ACC

-	(1790)	(1883)	(20)	
5	5	Ĩ.	(2)	
	559 560 ays 1ys TGC AAG	AAA	¥11.	
	559 973 160	T te	1101	
	558 675 100	Sc	AAAT	
10	1ys AAG	₹,	ACAT	
••	976	ter .	AAA	•
	val GTA	Jeu TTA	TCTA	
	ag t	\$ 55 56	crc	
	ele GCT	16. 17.)CAC	
15	ele cct	۽اء درد	√ 225	
	ohe TTC	213 GCT	LYYY	
	550 559 560 glu gin ieu iya ala val met asp aap ohe ala ala ohe val giu iya cys oys iya GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT CCT TTT GTA GAG AAG TGC TGC AAG	570 phe ala glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter TTT GCC GAG GAT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA CATCACATTTAAAAG (AAAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTTTT	6
	8.8 5.47	ser AGT	CGTT	(207
	met ATG	818 CCA	111	₹.
20	va1 GTT	al a	T1CT	:
	ele cct	val	GT T T	92
	£ \$	leu CTT	ATCT	:
	Je CTG	178 AA	ATTC	TA.
25	a to	1ya	CCTT	₩C
	91¢ 646	aly GCT	WYY.	AAG
	173 AA	9 2	GATC	ATGG
	540 1ys ala thr lys g AAG GCA ACA AAA G	570 91u GAG	Š	¥
	ala GCA	ala 000	¥	MTA
30	173 AAG	phe TTT	¥¥€	AATT
	614 000	567 cys 1GC	, <u>5</u>	2110
	1ys AAG	thr	ter ter ITGAGAATAA	7070
	» Is	glu	16.46 76.46	515
	531 giu leu val lys his lys pro GAG CTC GTG AAA CAC AGG CCC	567 asp asp lys glu thr cys GAC GAT AAG GAG ACC TGC	ter ter Catctcagcctagcatgagataagag	TCATTITGCCTCTTTTCTCTGTGCTTCAATTAATAAAATGGAAAGAATCTAA 20AA (2078)
35	va1 GTG	asp GAT)CCT/	3001
	leu CTC	asp GAC	TCAC	1111
	531 970 646	561 ala GCT	CATC	TCA1

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

	, 5175 115011651166	sequence	15
		(30)	
5		AGC	
		ohe TTT	
		<u> </u>	
		phe TTT	
10		. CT	
		7 5 5	
		4 361 TC	
		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
15		. E .	
15		7 2 7	
		2 3 C	
		lys t	
		-18 p r o' -10 Met lys trp val tlu phe lle ser leu bee leu bhe ser GCTTTTCTCTTCTGCCACACCCCTTTGGCACA ATG AAG TGG GTA ACC TTT ATT TCC CTT CTT TTT CTC TTT AGC	
20		Y C	
	•	1666	
		נככנו	
		ACAG	
25		וככככ	
		STCA	
		77.07.0	7
		ירכדכ	
30		хтт	٠.
-		J	9 :
			φ
			- 5
			, ,
35			•

ser ale tyr ser ang gly val phe ang and ICG CCT TAT ICC AGG GGT GTG TTT CGT CGA

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

5	(170)	(1961)	(350)	(460)	(330)	(620)	(710)	230 alu GAA (300)
	20 173 AAA	8	80 Jeu	110 pro CCA	140 try TAT		200 cys 7CT	230 910 GAA
	phe TTC	Dhe TTT	thr	AAC	leu 17A	169 170 cys gln TGC CAA		۶ کی در چ
	asn AAT	a]r CAA	818 GCA	88 57	TAC	168 cys TGT	leu lys CTC AAG	phe TTT
10	o ja CAA	thr	ral GTT		1ys AAA	₽ ₹		alu GAG
10	9 te G&A	val	th.	1ys AAA	glu glu thr phe leu lys lys GAA GAG ACA TTT TTG AAA AAA	thr ACA	ala lys aln ara GCC AAA CAG AGA	phe pro lys ala qlu phe TTT CCC AAA GCT GAG TTT
	phe lys asp leu aly TTT AAA GAT FFG GGA	54 A	75 978 TCC	his CAC	leu lys TTG AAA	. by	1 × ×	\$ ₹
	leu aly TTG GGA	asn AAT	Jeo TTA	g In CAA	phe	ala GCT	313	pro
	asb CAT	val GTG	1ys AAA	Jeu TTG	thr	ala GCT	101	phe 1
15	. .	leu val TTA GTG	asp GAC	bhe TTC	950		ser	8 4 5 A C A
		40 val lys leu GTA AAA TTA	91y CGA	101 633	o ye	tyr lys Tat aaa	190 lys sla ser AAG GCT TCG	gin arg CAG AGA
	10 arg	to Val	70 phe 711	5 4 5 5 4 5	130 asn AAT		190 AG	220 367 AGC
	h1s CAT	his CAT	70 leu phe CTT TTT	asn AAT	asp GAC	160 ala lys arq GCT AAA AGG	y 1 e 200	leu CTG
	al a GCT	asp GAT	thr ACC	AGA	r ts	ala GCT	alu aly GAA GGG	313
20	lys ser glu val AAG AGT GAG GTT	ole GAA	62 oys asp lys ser leu his thr TGT GAC AAA TCA CTT CAT ACC	<u> </u>	phe	ohe TTT	asp	ala ang GCT CGC
	lys ser glu val AAG AGT GAG GTT	phe	leu CTT		8 8 GCT	phe TTC	67.6 000	
	aer AGT	34 cys pro phe TGT CCA TTT	ser TCA	glu pro GAA CCT	thr	leu phë CTT TTC		ela val CCA GTA
	1ys AAG	34 cys TGT	asp lys ser leu GAC AAA TCA CTT	e Ay			g}u GAA	
25	ele his GCA CAC	gln CAG	asp GAC	1ys AAA	124 met cys ATG TGC	pro glu leu CCG GAA CTC	asp	lys ala trp AAA GCA TGG
25	91. GCA	gln CAG	62 0y s TGT	8 3 CCA	val GTG	200		lys AAA
	-t t arg asp cca cat	leu CTT	60 glu asn GAA AAT	91 Cys TCT	asp GAT	ala OCC	lys leu AAG CTC	ag TTC
	+- 600	30 tyr TAT		90 0ys TGC	120 val GTT	150 tyr 1AT	180 Pro CCA	210 ala GCT
	arg CGT	914 CAG	ala CCT	asp GAC	g)u GAG	phe TTT	leu 17G	S YOY
30	r o ol phe arg	phe ala	glu ser		2 5	tyr TAC	5 C	glu arg GAA AGA
	, (16,	phe TTT	g or cyc	glu met GAA ATG	ACA ACA	pro CCT	177 973 1GC	917
	91.	နို့ ပိ	ale asp GCT GAT	gly glu met ala GGT GAA ATG GCT	leu val arg p TTG GTG AGA G	£13	177 1 818 818 Cys 1 1 GCT GCC TGC C	phe
	9-9-19 ACG	lle ATT	al a CCT	gly CCT	leu 776	5 Y	ala GCT	<u>\$</u>
		1ec 11G	val		87.9 VCA	AG 43	\$ \$	# X
35		va1 GTG	53 cys val TCT CTT	thr tyr ACC TAT	pro arg	200 818	asp lys GAT AAA	15 S
		leu TTG	thr	€ 8	leu r	lie ala arg arg his pro ATT CCC ACA AGA CAT CCT	ala GCT	ser 1
		21 818 GCC	173 AA	arg ccr	111 asn AAC	141 glu 11e CAA ATT	171 ala ala GCT GCT	201 ala ser leu gin lys phe gly GCC AGT CTC CAA AAA TTT GCA

10 \$\frac{\pi}{15}	y asp leu leu glu cys ala asp asp arq ala asp leu A GAT CTG CTT GAA TGT GCT GAT GAC AGG GCG GAC CTT	279 280 cys qlu lys pro leu leu qlu lys ser his cys ile IGT GAA AAA CCT CTG TTG GAA AAA TCC CAC TGC ATT (980)	310 phe val glu ser lys asp val cys lys asm tyr ala TTT GTT GAA AGC GAT GTT TGC AAA AAC TAT GCT (1070)	340 asp tyr ser val val leu leu arg leu ala GAT TAC TCT GTC GTG CTG CTG AGA CTT GCC (1140)	370 tyr ala lys val phe asp glu phe lys pro leu TAT GCC AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)	400 glu tyr lys phe gin asn ala leu leu val ard GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT CGT (1340)	437 b38 b40 ser arg asm leu qîy lys val qîy ser lys cys cys lys his TCA AGA AAC CTA GGA AAA GTG GGC AGC AAA TGT TGT AAA CAT (1430)	460 461 Jeu cys val leu his giu lys thr pro val ser TTA TGT GTG TTG CAT GAG AAA ACG CCA GTA AGT (1520)	500 su glu val asp alu thr tyr val pro lys G GAA GTC GAT.GAA ACA TAC GTT CCC AAA (1610)	530 asp lie cys thr leu ser glu lys glu arg gln lie lys lys gln thr ala leu val GAT ATA TGC ACA CTT TCT GAG AAG GAG AGA CAA ATC AAG AAA CAA ACT GCA CTT GTT (1700)
និ 15 .	asp leu leu glu cys ala asp asp ara ala GAT CTG CTT GAA TGT CCT GAT GAC AGG CCG	280 glu lys pro leu leu plu lys ser his cys GAA AAA CCT CTG TTG GAA AAA TCC CAC TGC	310 val glu ser lys asp val cys lys asn tyr GTT GAA AGT AAG GAT GTT TGC AAA AAC TAT	tyr ser val val leu leu leu arg leu TAC-TCT GTC GTG CTG CTG AGA CTT	lys val phe asp glu phe lys pro AAA GTG TTC GAT GAA TTT AAA CCT	lys phe gin asn ala leu leu val AAA TTC CAG AAT GCG CTG TTA GTT	437 438 aly ser lys cys cys lys GGC AGC AAA TGT TGT AAA	pro val CCA GTA	glu val asp glu thr tyr val pro GAA GTC GAT.GAA ACA TAC GTT CCC	53 g gin lie lys lys gin thr ala leu va A CAA ATC AAG AAA CAA ACT GCA CTT GT
និ 15 .	asp leu leu glu cys ala asp asp ara ala GAT CTG CTT GAA TGT CCT GAT GAC AGG CCG	280 glu lys pro leu leu plu lys ser his GAA AAA CCT CTG TTG GAA AAA TCF CAC	310 val glu ser GTT GAA AGT	tyr TAC	lys val phe asp glu phe lys AAA GTG TTC GAT GAA TTT AAA	lys phe gin asn ala leu leu AAA TTC CAG AAT CCG CTG TTA	437 438 aly ser lys cys cys GGC AGC AAA TGT TGT		glu val asp alu thr tyr val GAA GTC GAT. GAA ACA TAC GTT	g gin lie lys lys gin thr als le A CAA ATC AAG AAA CAA ACT GCA CT
និ 15 .	asp leu leu glu cys ala asp asp arg GAT CTG CTT GAA TGT GCT GAT GAC AGG	280 glu lys pro leu leu glu lys ser GAA AAA CCT CTG TTG GAA AAA TCC	310 val glu ser GTT GAA AGT	tyr TAC	lys val phe asp glu phe AAA GTG TTC GAT GAA TTT	lys phe gin asn ala leu AAA TTC CAG AAT GCG CTG	437 aly ser lys cys GGC AGC AAA TGT		glu val asp alu thr tyr GAA GTC GAT. GAA ACA TAC	o oin lie lys lys oin thr al A CAA ATC AAG AAA CAA ACT GC
និ 15 .	asp leu leu glu cys ala GAT CTG CTT GAA TGT GCT	280 glu lys pro leu leu glu lys GAA AAA CCT CTG TTG GAA AAA	310 val glu ser GTT GAA AGT	tyr TAC	lys val phe asp glu AAA GTG TTC GAT GAA	lys phe gin asn ala AAA TTC CAG AAT GCG	aly ser CCC ACC	st 's val leu his glu lys th it GTG TTG CAT GAG AAA AC	glu val asp glu thr GAA GTC GAT. GAA ACA	g gin lie lys lys gin thi A CAA ATC AAG AAA CAA ACI
និ 15 .	asp leu leu glu cys ala GAT CTG CTT GAA TGT GCT	280 elu lys pro GAA AAA CCT	310 val glu ser GTT GAA AGT	tyr TAC	lys val phe asp AAA GTG TTC GAT	lys phe gin asn AAA TTC CAG AAT	aly ser CCC ACC	st 's val leu his qiu ly it GTG TTG CAT GAG AA	glu val asp glu GAA GTC GAT. GAA	g ain lie iys lys air A CAA ATC AAG AAA CA
និ 15 .	asp leu leu glu cys ala GAT CTG CTT GAA TGT GCT	280 elu lys pro GAA AAA CCT	310 val glu ser GTT CAA AGT	tyr TAC	1ys AAA	lys phe AAA TTC	ly lys val qly se ga aaa gtg ggg ag	st 19 val leu his gl 17 GTG TTG CAT GA	glu val asp GAA GTC GAT.	g ain lie iys ly: A CAA ATC AAG AAV
និ 15 .	asp leu leu glu cys GAT CTG CTT GAA TGT	280 elu lys pro GAA AAA CCT	310 val glu ser GTT CAA AGT	tyr TAC	1ys AAA	lys phe AAA TTC	ly lys val al CA AAA GTG GG	51 's val leu hi it GTG TTG CA	glu	g ain lie lys A CAA ATC AAG
15 .	asp leu leu glu GAT CTG CTT GAA	280 elu lys pro GAA AAA CCT	310 val glu ser GTT CAA AGT	tyr TAC	1ys AAA	tyr lys pl	ly lys va EA AAA GI	51 'S val le IT GTG TT	glu	a aln 116 A CAA ATG
	asp leu leu GAT CTG CTT	280 91u 1ys GAA AAA	310 val glu GTT GAA	tyr TAC	ala I	tyr 13	 	33 34 5 8		53
250	83D CAT	280 41 644	310 val GTT	340 15p t	# O	<i>E</i> 2			F	
	83D CAT	279 ; cys TGT (•• • • • • • • • • • • • • • • • • • •		370 tyr TAT	\$00 910 100	436 1ev a CTA G	460 461 leu cys TTA TGT	490 ala leu GCT CTG	A 20
			- €	pro d	369 3 cys tr TGC T	9 6 60 60 4 9	£ 5	2 1 6 G		520 9 4 Lu 6 AG
	6 8	278 0ys 1CC	asp p	AT C	369 his alu cys CAT GAA TGC	leu al CTT G	# ¥ 6 ≾	n gln c cAG	phe ser TTT TCA	
20	his gly cat goa	2,3	ala a	arg his AGG CAT	his a	aln leu CAG CTT	# * *	u asn G AAC		29
246		278 Ser ser lys leu lys glu oys TCC AGT AAA CTG AAG GAA TGC		ala arg arg his GCA AGA AGG CAT	pro h		val se GTC TC	tyr leu ser val val leu asn TAT CTA TCC GTG GTC CTG AAC	pro cys CCA TGC	<u>.</u> 5
245	glu cys cys GAA TGC TGC	513	lev pro ser leu ala TTG CCT TCA TTA GCT	ala arg GCA AGA	ala ala asp pro GCT GCA GAT CCT	phe alu TTT CAC	glu val GAG GTC	l val c crc		<u> </u>
	28	* X	ser leu TCA TTA	tyr a	ala a CCA G			r val C GTG	a arg	\$ £ \$
25		150	5 5	glu tyr GAA TAT	ala a GCT G	glu leu GAG CTT	leu val CTT GTA	u ser A TCC	arg c AGG	514 cys
25	lys val his thr AAA GTC CAC ACG	- 20 - 20 - 20 - 20 - 20 - 20 - 20 - 20	300 ala asp leu pro GCT GAC TTG CCT	tyr g	ala a OCC O	392 cys g TGT G	thr leu val ACT CTT GTA	tyr leu TAT CTA	l asn G AAC	4 11
,	CTC .	11e ATC	889 J	leu t 776 7	361 cys a TGT 0		pro th	7€	u val	S S S
240	lys AA	270 9er 100 /	300 818 6CT 0	330 phe 1 TTT T	360 3 cys c TCC T	390 gln asn CAA AAT	420 thr pro ACT CCA	450 glu asp GAA GAC	480 ser leu TCC TTG	510 his ala cat cca
	thr Acc /	esp s	pro #	3 met p ATG T	3 1ys c; AAG T(173 91 AA C	420 r thr A ACT		480 3er	510 his
30	leu thr CTT ACC	£ ₹	met p	91y # CCC A	glu 13 GAG A		1 ser G TCA	8 s ala T GCA	. 9lu CAA	phe
			glu met CAG ATG			. ∓.₹ ≳.g	n val A GTG	448 0 cys C 1CT	thr ACA	thr
	val thr asp GTG ACA GAT	265 ays glu esn TGT GAA AAT	asp glu GAT GAG	phe leu TTC TTG	glu thr thr leu GAA ACC ACT CTA	asn leu ile AAT TTA ATC	pro gln ccc cAA	met pro ATG CCC	476 477 Cys Cys TCC TCC	af T
,	16 / DE	265 9ys 9 1GT 0	8 ± C		. F O			met ATG	476 cys TGC	AC S
	2 € 7 €	1e o TC T	# ? ≥ \$	asp val GAT GTC	ه د ک	pro gln CCT CAG	8 val	919	val thr lys GTC ACC AAA	gle GAA
35	lys leu AAG TTA	tyr lle TAT ATC	2 5 2 9	% C3	22	20	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	173	thr	ala CCT
,	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	78 2 5	2 X	₽ <u>₹</u>	r tyr A TAT	glu glu GAA GAG	1ys	giu ala iya arg GAA GCA AAA AGA	val GTC	asn AAT
	val se	261 ala lys GCC AAG	291 ala glu val glu asn GCC GAA GTG GAA AAT	321 glu ala lys GAG GCA AAG	351 1ys thr tyr glu thr AAG ACA TAT GAA ACC	381 val glu glu pro gln GTG GAA GAG CCT CAG	411 tyr thr lys lys val TAC ACC AAG AAA GTA	9 4 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	471 88p 8rg GAC AGA	501 glu phe asn ala glu thr phe CAG TIT AAT GCT GAA ACA TTC
~	3 6	พีซีซี	8 6 8	321 91u GAG	351 173 AAG	381 val GTG	411 tyr 1AC	641 070 CCT	471 38p CAC	501 glu GAG

5	559 560 cys lys TGC AAG (1790)	ter TTAAAAG (1883)	TCTTTAA (2002)
10	558 559 5 phe val glu lys cys cys l TTT GTA GAG AAG TGC FGC A	ter TTA GCC TTA TAA CATCACATTTAAAAG (1883)	CTGTCTAAAAACATAAAT!
15	550 asp asp phe ala ala phe val glu l CAT GAT TTC GCT GCT TTT GTA GAG A	580 gin ala ala leu qiy r CAA GCT GCC TTA GCC	rgetgtaaagccaacacc 78)
20	540 1ys ela thr lys giu gin leu lys ele val met esp AAG GCA ACA AAA GAG CAA CTG AAA GCT GTT ATG GAT	580 qly lys lys leu val ala ala ser gin ala ala i GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC	VTCTGTTTTTCTTTTTCGT1
25	0 r lys glu gin leu A AAA GAG CAA CTG	0 u glu qly lys lys G-GAG GGT AAA AAA	AGATCAAAAGCTTATTCA AATGGAAAGAATCTAA
30	5 TO CO	567 570 thr cys phe ala glu glu q ACC TGC TTT GCC GAG-GAG G	ter Ataagagaagaaatga Gegettgattaataaa
35	531 glu leu val lys his lys GAG CTC GTG AAA CAC AAG	561 ala asp asp lys glu thr cys GCT GAC GAT AAG GAG ACC TGC	ter ter Catetcagcetaccatgagaataagaaaatgaagatcaaaagcttattcatetgtttttegttggtgtaaagccaacaccetgtetaaaacataaatttetttaa (2002) TCattitgcetettttetetgtgettgaataaaaatggaaagaatetaa 20aa (2078)

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

5	(30)	(170)	(192)	(350)	(660)	(330)	(620)	(710)	(300)
	A Se	20 1ys AAA	5 s s 5	80 CTT	110 070 CCA	140 try TAT	170 61n CAA (200 cys TGT (230 alu GAA (
	4 F	phe TTC	phe	thr	asn AAC	leu TTA	169 cys TG:	1ys AG	818 GCA
	182 CTC	AAT	alu CAA	\$18 CCA	350	1 yr	168 cys 1GT	leu CTC	
10	phe TTT	alu GAA	thr	val	aso CAT	173 AA	alu GAA	r S	alu phe GAG TTT
10		2 ₹		th ACA		lys AA	thr ACA	ala CAG /	ala GCT C
	2 E	ور ور خ	glu val GAA GTA		his lys Cac aaa	leu lys TTG AAA	phe TTT	lys aln AAA CAG	pro lys ala alu phe CCC AAA GCT GAG TTT
	100	leu 11G	AAT	75 leu cys TTA TCC	GAA CAA	phe	ala ccr	818 GCC	pro 1
	11e	SAT	val GTG		1 Tec	thr ACA	ala d	3er 8	phe p
15	ag E	17. A.A.	leu val TTA GTG	asp lys GAC AAA	ohe TTC	glu thr GAG ACA	tyr lys ala TAT AAA GCT		ACA T
	AG the		Tys AAA	4 2	101 698 100 100	alu GAA	tyr 1	190 lys ala ser AAG GCT TCG	220 ser gin arg AGC CAG AGA
	p r c trp val tlu TGG GTA ACC	10 arg phe CCC TTT	40 val lys GTA AAA	Dhe TTT	100 101 glu eys GAA TGC	130 asm qlu AAT GAA	160 arg 1 Acc 1	190 173 AG G	220 ser gin AGC CAG
	t 5	CAT S	N.S.	Jeu CTT	AAT	ase GAC	lys a		15 Z
		SCT CCT	SSP CAT	A thr	P S	his a	ala lys CCT AAA	alu aly GAA GGG	ala arg leu GCT CGC CTG
20	-18 Wet 1ye ATG AAG		88) of 6	phe 1	ohe a	885 GAT C	ala a
		glu val GAG GTT	phe TT	leu his CTT CAT	679	sis r	phe p TTC 1	6.00	
		AGT	01 A	ser ICA (9 to 649	thr a		le CTT C	ala val GCA GTA
		1ys	¥ 875	173 AA :	gin glu CAA GAA	124 073 100 /	leu l	glu leu GAA CTT	trp a
25		41. CAC	gln CAG	62 asn cys asp lys ser leu his AAT TGT GAC AAA TCA CTT CAT	1 ys	ATG 1	glu leu leu GAA CTC CTT	asp g	ala t CCA 1
LJ		ala his cca cac	gla CAG	62 cys asp 1GT GAC	818 1 CCA 1	val .	0 010 0 000	leu a	lys a
		CAT C	Jeu c	asn c	91 Cys 8	asp v	ala p	lys l	를 보고 무 도
			30 tyr 1 TAT 0	8 26 8	90 91 cys cys sle lys rcc rcr cca AAA	120 val a GTT 0	150 tyr a	180 pro lys CCA AAG	210 ala phe GCT TTC
		phe arg arg TTT CGT CGA	gln t	ala g	asp c	120 glu val GAG GTT	phe t	leu p	arg a
30		7 T T T	ala CCT C	36. 77.	ala CCT C	2 Y	tyr p TAC T	leu l	8 248 8 4 89
			phe a		_			~ % S	91y 9 CGA G
		ply val	ala 000	asp glu GAT GAG	glu met GAA ATG	arg leu val arg CCA TTG GTG AGA	his pro CAT CCT	177 ala eys GCC TGC	he T G
		9-10 9-00 9-00	ile a	31.8 SCT 0	aly 9 ccr c	5 G > G	5 Y	ala a	gin lys phe c
			1eu -	val a	tyr g TAT G	53	E &	lys ala	
35		× × ×	val l	53 cys val TGT GTT	thr ty ACC T	6 00 000 000 000 000 000 000 000 000 00	ala arg arg OCC AGA AGA		16 9 17 9
		als tyr ser GCT TAT TCC	leu val TTC CTC	thr c	gle ti	leu pi CTC C	11e al Att C	ala asp GCT GAT	# 5 # #
		# D	21 ala 16 CCC T	S1 lys thr AAA ACA					201 ala ser GCC AGT
		= =	<u></u> 8	~ ~ ₹	81 819 CCT	111 asn AAC	돌물중	171 818 GCT	201 818 GCC

5	(890)	(980)	(1070)	(1160)	(1250)	(13%0)	(1430)	(1520)	(1610)	(1700)
!	2&0 1eu CTT	290 13e ATT	320 ala CC:T	3Šn 818 GCC	380 162 CCT	410 810 CGT	880 N13	470 ser AGT	Sno 1ys AAA	530 val
10	35	289 cys TGC	tyr TAT	Jeu CTT	oro CCT	val GTT	1ys AAA	val GTA	در درر	leu CTT
	a 1 a CC	h is CAC	AAC	ACA ACA	1ys AAA	leu TTA	437 438 cys cys TGT TGT	878 CCA		818 GCA
	P CC	36T	lys AAA	leu CTG	phe TT	Jeu CTG	437 cys TCT	thr Acc	tyr val IAC GIT	thr ala ACT GCA
10	SSD	glu lys GAA AAA	316 cys lys TGC AAA	leu leu CTG CTG	2 ₹	al a တင	lys AA	alu lys GAG ĀĀĀ	thr ACA	₽
	ala asp asp and ala GCT GAT GAC AGG GCG	£ X	316 lys asp val cys lys asm AAG GAT GTT TGC AAA AAC	ser val val leu leu leu ard TCT GTC GTG CTG CTG AGA	asb GAT	40A glu tyr lys phe gln asn ala leu leu val GAG TAC AAA TTC CAG AAT GCG CTG TTA GTT	AGC		490 ser ala leu glu val asp glu thr TCA GCT CTG GAA GTC GAT GAA ACA	520 glu arq ain ile lys lys ain thr ala GAG AGA CAA ATC AAG AAA CAA ACT GCA
	s s CCT	leu leu CTG TTG	asp CAT	val	phe TTC	16. CAG	val qly GTG GGC	leu his TTG CAT	asp GAT	1ys AAG
	253 cys TGT	1ee CTG	1ys AAG	val	val GTG	age 71	val GTG		val	11e ATC
15	253 leu plu cys CTT GAA TGT	77 CCT	glu ser GAA AGT	367 TCT	ala lys GCC AAA	lys AA	1ys AAA	461 cys val TGT GTG	ale GAA	ag CA
	250 leu leu CTG CTT	1ys AA		tyr TAC	နူး ဗင္ဗ	tyr TAC	417 CGA	461 cys TGT	leu CTG	AGA
	250 CTG	280 410 644	310 asp phe val GAT TTT GTT	340 pro asp CCT GAT	369 370 pro hís qiu cys tyr CCT CAT GAA TGC TAT	400 91c	430 leu CTA	\$60 gIn Ieu CAG TTA	490 ala CCT	520 Leu ser glu lys glu CTT TCT GAG AAG GAG
	GAT CAT	278 279 ser lys leu lys glu cys cys AGT AAA CTG AAG GAA TGC TGT	a Phe	pro CCT	369 cys TGC	9. 56.4	AAC		phe ser TTT TCA	leu ser glu lys CTT TCT GAG AAG
20	91,000	278 cys 700	885 GAT	EAT CAT	g lu GAA	leu CTT	8 75 A54	AAC	phe TTT	g)u GAG
20	265 266 oys oys his gly TGC TGC CAT GGA	3 to 3	ala ala GCT GCT	glu tyr ala arg arg his GAA TAT GCA AGA AGG CAT	pro his qiu CCT CAT GAA	glu gln leu GAG CAG CTT	ser TCA	ser val val leu asn TCC GTG GTC CTG AAC	pro cys phe CCA TGC TTT	ser TCT
	246 973 100	lys AG	s la GCT	arg AGA	pro CCT	a]r	val GTC	val GTC	77 CCA	leu CTT
	265 glu oys GAA TGC	leu CTG	ser leu TCA TTA	هاه 67۸	ala ala ala asp. GCC GCT GCA GAT	phe TTT	glu	val GTG	leu val asn arg arg TTG GTG AAC AGG CGA	514 gys thr TGC ACA
		1ys AAA		tyr TAT	\$1\$ CCA	glu leu GAG CTT	thr leu val ACT CTT GTA	36T	879 AGG	510 his ale asp lie cys CAT GCA GAT ATA TGC
25	hie thr	ser AGT	leu pro TTG CCT	g lu GAA	ala ala GCC GCT		thr leu ACT CTT	tyr leu TAT CTA	asn AAC	510 his ale asp ile CAT GCA GAT ATA
		ser TCC	leu TTG	tyr TAT		392 cys 1GT	thr	tyr 1AT	val GTG	asp GAT
	240 lye vel AAA GTC	270 ser 11e 7CG ATC	300 ala asp GCT GAC	leu 11G	360 361 0ys cys TGC TGT	390 gln asn CAA AAT	420 thr pro ACT CCA	450 glu esp GAA GAC	1eu	ala GCA
				330 phe 111					480 3er 100	
	th Acc	gin asp CAA GAT	pro CCT	gly met GGC ATG	glu lys GAG AAG	ile lys ATC AAA	ser TCA	448 cys ala TGT GCA	g lu SA	thr phe ACC TTC
30	ST	gla	met ATG	91.y CCC		•	val		thr	-
	8. S.	265 cys glu asn TGT GAA AAT	9 te	Jeu TTG	leu CTA	gin asn leu CAG AAT TTA	914 CAA	979	477 cys TGC	phe TTC
	thr ACA	g Po	es CAT	phe TTC	thr	gin asn CAG AAT	. 5 00	met ATG	476 lys cys	thr ACA
	** CTG	265 cys 1GT	gtu asn GAA AAT	val GTC	thr		val GTA	arg AGA	1ys	95 84
35	leu 17	tyr lle TAT ATC	glu	asp GAT	ole AA	glu pro GAG CCT	1ys 1ys AAG AAA	1 ys	val thr GTC ACC	al a
	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	tyr TAT	glu val GAA GTG	ala lys asp GCA AAG GAT	tyr	glu GAG	thr lys lys val pro ACC AAG AAA GTA CCC	glu ala lys arg met CAA GCA AAA AGA ATG	va1 570	asn AAT
		261 ala lys tyr lle GCC AAG TAT ATC		al a GCA	351 1ys thr tyr glu thr thr leu AAG ACA TAT GAA ACC ACT CTA	381 val glu glu pro GrG GAA GAG CCT	thr	8 8	471 asp arg val thr GAC AGA GTC ACC	501 glu phe asn ala glu thr phe GAG TIT AAT GCT GAA ACA TTC
	23. 16.7 CT	261 ala GCC	291 818 GCC	321 91u GAG	351 1ys AAG	381 val GTG	411 tyr TAC	pro CCT	471 889 GAC	501 91u GAG

5	559 560 cys lys TGC AAG (1790)	AAG (1883)	TAA (2002)	
10	540 Sys ale thr lys glu gin leu lys ele val met esp asp phe ale ale phe vel glu lys cys lys AMC CCA ACA AMA GAC CAA CTC TOT ATG GAT GAT TTC GCT TTT GTA GAG AGG TGC TGC AAG	570 glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GOC TTA TAA CATCACATITAAAAG (1883)	AGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAAACATAAATTTCTTTAA (2002) NTAATAAAAAATGGAAACAATGTAA	
15	550 GAT TTC GCT GCT TTT	o n ala ala leu qly A CCT CCC TTA CCC	GTAAAGCCAACACCCTGT	
20	55 ale val met asp as CCT GTT ATG GAT GA	580 val ala ala ser gin GTT GCT GCA AGT CAA	GTTTTCTTTTCGTTGGT	
25	Iys glu gin leu iys AAA GAG CAA CTG AAA	glu qly lys lys leu GAG GGT AAA AAA CTT	ATCAAAGCTTATTCATCT TGGAAAGAATCTAA	
30		567 cys phe als TGC TTT GCC		
35	531 glu leu val lys his lys pro GAG CTC GTG AAA CAC AAG CCC	561 ala asp asp lys glu thr GCT GAC GAT AAG GAG ACC	ter ter Catctcagcctagcatgagataagaaaatgaagatcaaaagcttattcatctgtttttgtttg	1
		_		

- 10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
- 11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
 - 13. A DNA transfer vector according to claim 12, which is a plasmid.
- 14. A DNA transfer vector according to claim 13,10 wherein the plasmid is pBR322 or YEp6.
 - 15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
- 16. A DNA transfer vector according to any of 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
 - 17. A vector or process according to claim 16, wherein the bacterium or yeast is <u>E. coli</u> or <u>Saccharomyces</u> <u>cerevisiae</u>.

1/1 = (8) 1 (8) 2.0 586 form TAA Sst | 531/2 Restriction Endonuclease Map of Human Serum Albumin cDNA Clones 493/4 Tag 1 Hinf 1 479/0 419/0 450/1 Hinc II Mbo II **pHA36** 382 Mba 12 Kilobases 325/6 Mbo 11 6. Hinf 1 269/0 Pat 1 (3611) Hpa II (3548) 182/3 Tag ! œ 4 Hpa II (3548) pHA206 57 Hinf 1 Pst 1 Mbo 11 Mbo 11 (3611) 16/7 31 Hpa II (3658)